week8

krishna sai surendra babu kalluri

01/08/2020

1)Here,we created a binary variable for Auto dataset. we wrote a condition as per the requirement.we can see the output that it satisfies the condition.

```
knitr::opts_chunk$set(echo = TRUE)
require(ISLR);

## Loading required package: ISLR

require(tidyverse);

## Loading required package: tidyverse

## -- Attaching packages ---------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.0
## v tidyr   1.1.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts ------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

require(ggthemes)

## Loading required package: ggthemes

require(e1071);

## Loading required package: e1071

require(caret)

## Loading required package: caret

## Loading required package: lattice

##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
##     lift
```

```
theme_set(theme_tufte(base_size = 14))
set.seed(1)

data(Auto)
Auto <- as.tibble(Auto)

## Warning: `as.tibble()` is deprecated as of tibble 2.0.0.
## Please use `as_tibble()` instead.
## The signature and semantics have changed, see `?as_tibble`.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.

Auto <- Auto %>%
    mutate(highmpg = as.integer(mpg > median(mpg))) %>%
    mutate(highmpg = factor(highmpg),
           cylinders = factor(cylinders))

Auto %>%
    sample_n(10) %>%
    select(mpg, highmpg)

## # A tibble: 10 x 2
##      mpg highmpg
##    <dbl> <fct>
##  1  44.3 1
##  2  23   1
##  3  26   1
##  4  23.9 1
##  5  23.2 1
##  6  16   0
##  7  33.5 1
##  8  13   0
##  9  31.5 1
## 10  17.6 0

Auto <- Auto %>%
    select(-mpg,-name)

some_var <- dummyVars(highmpg ~ ., data = Auto)
test_var <- predict(some_var, Auto)

## Warning in model.frame.default(Terms, newdata, na.action = na.action, xlev
=
## object$lvls): variable 'highmpg' is not a factor
```

#a)

```
knitr::opts_chunk$set(echo = TRUE)
linear_model <- train(x = test_var, y = Auto$highmpg,
                      method = 'svmLinear2',
                      trControl = trainControl(method = 'cv', number = 10,
allowParallel = TRUE),
                    preProcess = c('center', 'scale'),
                    tuneGrid = expand.grid(cost = c(1,5,7.5,15)))

linear_model$finalModel

##
## Call:
## svm.default(x = as.matrix(x), y = y, kernel = "linear", cost = param$cost,
##     probability = classProbs)
##
##
## Parameters:
##    SVM-Type:  C-classification
##  SVM-Kernel:  linear
##        cost:  15
##
## Number of Support Vectors:  74
```

A support classifier is fitted for our data for different cost values for linear model and error for different values.The cost value can be varied for our classifier ,cost argument will allow to specify the cost of violation for the margin.If the cost argument is small then the margins are wide ,many support vectors will be on the margin or it will violate the margin.If the cost is large then margins are narrow and few support vectors will be on the margin or violate the margin.Here,cost 1 is the best classifier

#b)

```
knitr::opts_chunk$set(echo = TRUE)
polynomial_model <- train(x = test_var, y = Auto$highmpg,
                  method = 'svmPoly',
                  trControl = trainControl(method = 'cv', number = 10,
allowParallel = TRUE),
                  preProcess = c('center', 'scale'),
                  tuneGrid = expand.grid(degree = c(2,3,4),C =
c(1,5,7.5,15),scale = TRUE))
polynomial_model$finalModel

## Support Vector Machine object of class "ksvm"
##
## SV type: C-svc  (classification)
##  parameter : cost C = 1
##
## Polynomial kernel function.
##  Hyperparameters : degree =  2  scale =  TRUE  offset =  1
##
## Number of Support Vectors : 71
##
## Objective Function Value : -45.587
## Training error : 0.045918
```

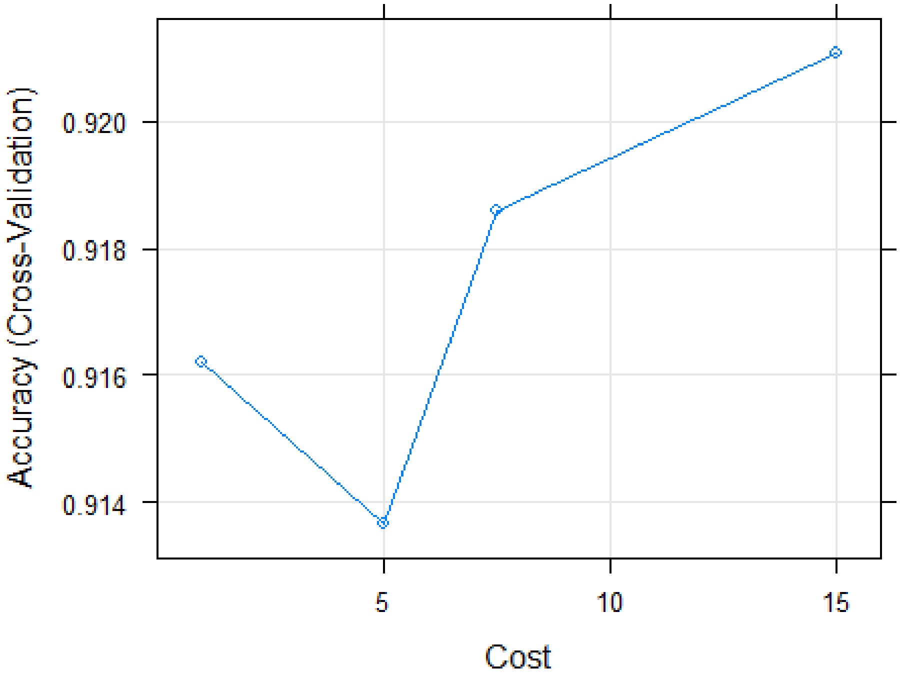
```
knitr::opts_chunk$set(echo = TRUE)

radial_model <- train(x = test_var, y = Auto$highmpg,method =
'svmRadial',trControl = trainControl(method = 'cv', number = 10,
allowParallel = TRUE),preProcess = c('center', 'scale'),tuneGrid =
expand.grid(C = c(1,5,7.5,15),sigma = seq(0.1, 6.5, length.out = 5)))
radial_model$finalModel
```

```
## Support Vector Machine object of class "ksvm"
##
## SV type: C-svc  (classification)
##  parameter : cost C = 1
##
## Gaussian Radial Basis kernel function.
##  Hyperparameter : sigma =  1.7
##
## Number of Support Vectors : 236
##
## Objective Function Value : -75.64
## Training error : 0.022959
```

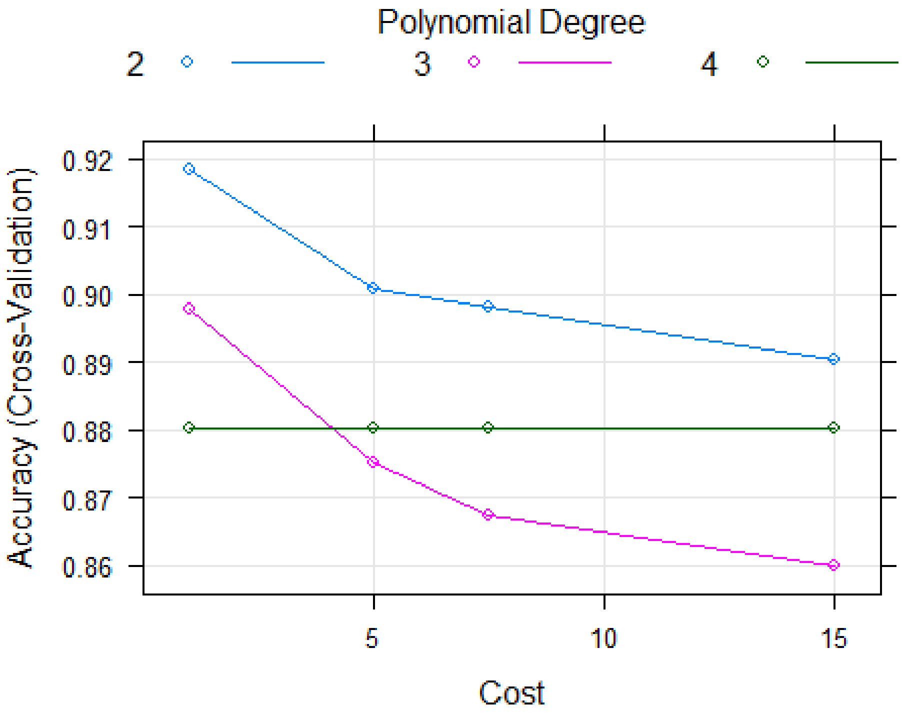
From the above code ,we can interpret that accuracy of dataset can change with kernels .By using Svm we can see a fair number of training errors.If we increase the cost , we can reduce the number of training errors.

#c)

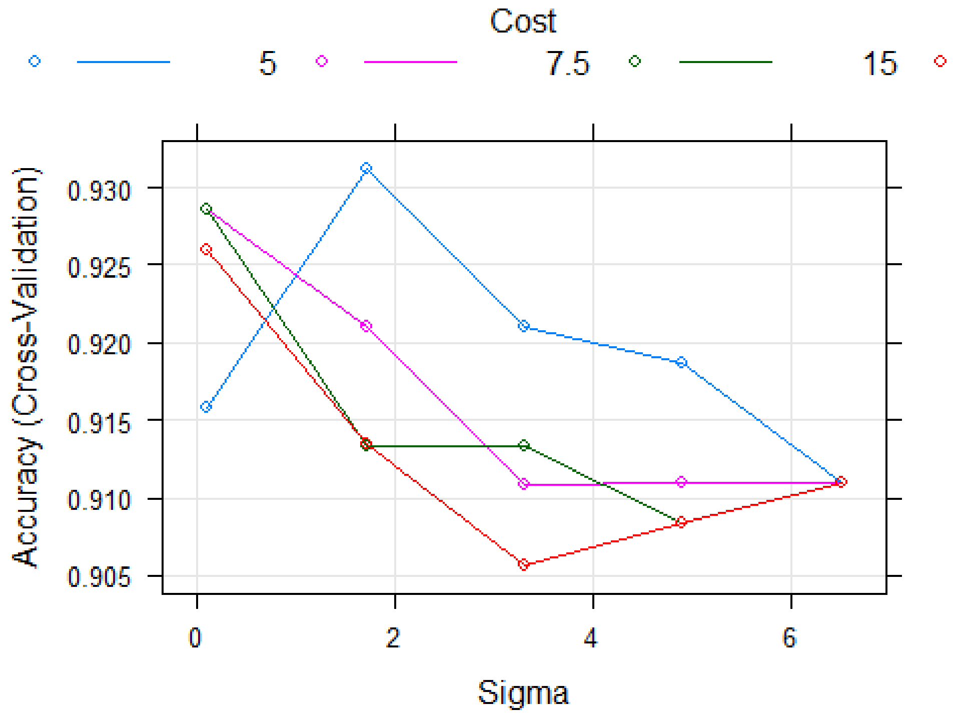
```
knitr::opts_chunk$set(echo = TRUE)

plot(linear_model)
```



```
plot(polynomial_model)
```



```
plot(radial_model)
```



#d)

```
knitr::opts_chunk$set(echo = TRUE)
postResample(predict(linear_model),Auto$highmpg)

##  Accuracy     Kappa
## 0.9311224 0.8622449

postResample(predict(polynomial_model),Auto$highmpg)

##  Accuracy     Kappa
## 0.9540816 0.9081633

postResample(predict(radial_model),Auto$highmpg)

##  Accuracy     Kappa
## 0.9770408 0.9540816
```

Here, we used postResample method in order to find out the accuracy for three models and Radial model is the best model with accuracy rate of 0.9770408

R Markdown

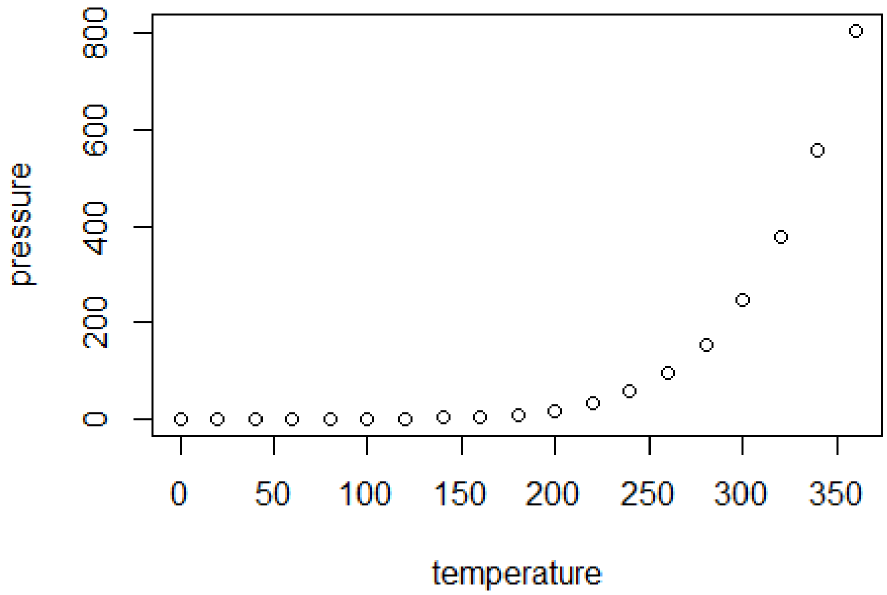
This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)

##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.